Active Learning with Ensembles for Image Classification

H. Liu and **A. Mandvikar**[1] and **P. Foschi**[2] and **K. Torkkola**[3]
[1]Department of Computer Science & Engineering, PO Box 875406,
Arizona State University, Tempe, AZ 85287. {huanliu,amitm}@asu.edu
[2]Romberg Tiburon Center, 3152 Paradise Drive, Tiburon, CA 94920. tfoschi@sfsu.edu
[3]Human Interface Lab, Motorola Labs,Tempe, AZ 85284. kari.torkkola@motorola.com

Abstract

In many real-world tasks of image classification, limited amounts of labeled data are available to train automatic classifiers. Consequently, extensive human expert involvement is required for verification. A novel solution is presented that makes use of active learning combined with an ensemble of classifiers for *each class*. The result is a significant reduction in required expert involvement for uncertain image region classification.

1 Introduction

Multimedia contents are rapidly becoming a major target for data mining research. This work is concerned with image mining, discovering patterns and knowledge from images for the purpose of classifying images. The specific problem we address is image region classification. *Egeria Densa* is an exotic submerged aquatic weed causing navigation and reservoir-pumping problems in the Sacramento-San Joaquin Delta of Northern California. As a part of a control program to manage Egeria, classification of regions in aerial images is required. This problem can be abstracted to one of classifying massive data without class labels. Relying on human experts for class labeling is not only time-consuming and costly, but also unreliable if the experts are overburdened with numerous minute and routine tasks. Massive manual classification becomes impractical when images are complex with many different objects (e.g., water, land, Egeria) under varying picture-taking conditions (e.g., deep water, sun glint). The main objective of the project is to relieve experts from going through all the images and pointing out locations where Egeria exists in the image. We aim to automate the process via active learning to limit expert involvement to decisions about which the automatic classifier is uncertain.

The contributions of this work are a novel concept of *class-specific ensembles*, and learning class-specific ensembles. We notice that different types of classifiers are better suited to detecting different objects such as Egeria, land, water. Since it is impractical to train one classifier for each object (as experts need to provide training instances for all objects), we propose a novel approach to class-specific ensembles. We also discuss how to combine individual classifiers to learn class-specific ensembles. We show that this approach significantly reduces the number of uncertain image regions and is better than a single ensemble for the task of object detection.

2 Class-Specific Ensembles

We may tend to use as many classifiers in an ensemble as possible because (1) each classification algorithm may have a different view of the training image and capture varied aspects of the image, as different classification algorithms have different biases and assumptions; and (2) no single classifier can cover all aspects. In other words, some algorithms may succeed in capturing some latent information about the domain, while others may fail. However, problems can result from using too many classification algorithms: e.g., (1) using more classification algorithms can result in longer overall training time, especially so if some of the algorithms are time-consuming to train; and (2) some algorithms may be prone to overfitting. If these algorithms are included in the ensemble, there may be a high risk of allowing the ensemble to overfit some training image(s). The above analysis suggests the necessity of searching for a relevant set of classifiers to form ensembles. Exhaustive search for the best combination is impractical because the search space is exponential in the total number of classification algorithms for consideration. Thus we need an efficient methodology to find the optimal combination of classifiers for the class-specific ensembles.

We adopt three performance criteria for this work: Precision, Recall, and number of uncertain regions (UC). The first two standard measures are defined in terms of the instances that are relevant and the instances that are classified (or retrieved). The third one is the number of instances (regions in an image) the class-specific ensembles cannot be certain about their class predictions. High precision or high recall alone is not a good performance measure as each describes only one aspect of classification. Together they provide a good measure: for example, the product of precision and recall (PR). If both values are 1, their product is 1, which means all and only positive instances are classified as positive. Hence, PR is a measure for both generality and accuracy. F measure can be used for the same purpose.

3 Learning Class-Specific Ensembles

In order to learn class-specific ensembles, we need a data set that links classifiers to the label of each image region. This

input: Tr, l_{Tr}, E : Set of n classification algorithms
output: $E_{l=yes}, E_{l=no}$

Train E with Tr to obtain n classifiers, cl_1 to cl_n;
Obtain class labels, l^1_{Tr} to l^n_{Tr} for Tr using cl_1 to cl_n;
Form a data set, $D \leftarrow \{l^1_{Tr}, l^2_{Tr}, ..., l^n_{Tr}, l_{Tr}\}$;
Learn association rules, $Assoc$ from D;
$Assoc_1 \leftarrow$ Filter ($Assoc$ / consequent is $l_{Tr} = yes$);
$Assoc_0 \leftarrow$ Filter ($Assoc$ / consequent is $l_{Tr} = no$);
$m_{l=yes} \leftarrow$ Max $(Assoc_1, supp * conf)$;
$m_{l=no} \leftarrow$ Max $(Assoc_0, supp * conf)$;
$Assoc_1 \leftarrow$ Filter $(Assoc_1 / supp * conf \geq 0.9 * m_{l=yes})$;
$Assoc_0 \leftarrow$ Filter $(Assoc_0 / supp * conf \geq 0.9 * m_{l=no})$;
$Assoc_1 \leftarrow$ Sort $(Assoc_1, length(precedent))$;
$Assoc_0 \leftarrow$ Sort $(Assoc_0, length(precedent))$;
$E_{l=yes} \leftarrow$ Precedent(First($Assoc_1$));
$E_{l=no} \leftarrow$ Precedent(First($Assoc_0$));

Figure 1: Algorithm for selecting optimal classifiers

new data set can be obtained by applying all the classification algorithms to the training image so that each classifier is a feature (i.e., a column) and its value is the prediction of the classifier. For each image region (one instance in the new data set), there are predictions of all the classfiers and also the class label 'Egeria' or 'non-Egeria' given by experts. We are concerned only with those rules that have the class label Egeria or not as the consequent. We will restrict our search to such rules and obtain rules with the maximum number of features (classifiers) in the precedent without a significant loss in support or confidence. The best rule for each class label indicates the best combination of classifiers for the ensemble.

Among many learning algorithms for classification, clustering, and association rules, we observe that association rule algorithms [Agrawal and Srikant, 1994] can search the attribute space to find the best combination of attributes associated with a class. Our algorithm is presented in Figure 1. It takes as input the entire set of classification algorithms E and training data T_r with class labels l_{Tr}, and produces as output two ensembles for class label *yes* and class label *no*.

The next task is to use the dual ensembles ($E_{l=yes}$ and $E_{l=no}$) to determine certain and uncertain instances. We need to decide the maximum number of classifiers in an ensemble that should agree on a prediction to reach a decision of "certain" or "uncertain" for each ensemble. An ensemble with all classifiers being required to agree on a prediction would lead to high precision, but low recall; an ensemble with few classifiers being required to agree would lead to high recall and low precision. We need to find a balanced number of classifiers with which the ensemble gives the best estimated precision and recall. The training image is used again for this task.

The dual ensembles $E_{l=yes}$ and $E_{l=no}$ work together to decide if an instance's prediction is certain or not following the rule of majority. In predicting an instance, if both $E_{l=yes}$ and $E_{l=no}$ are certain with their predictions, follow the one with more agreeing classifiers; if one is certain and the other is uncertain, follow the certain one; if both are uncertain, the instance is uncertain.

4 Experiments and Evaluations

With the principal goal of reducing the burden on experts, we use only one image for training and apply the learned results to another 16 testing images of different areas for detection. Among the classification algorithms available in the machine-learning package WEKA [Witten and Frank, 2000], we select all that can be applied to the image domain to ensure a variety of classification algorithms. We apply the algorithm in Figure 1 with the complete set of classification algorithms as input. The class-specific ensembles found by association rules are: **Class = yes** $\Leftarrow$ Decision Trees C4.5, Alternating Decision Trees, Decision Trees (PART), PRISM, Hyper Pipes, Kernel Density, Logistic, Decision Tables; and **Class = no** $\Leftarrow$ Id3, Alternating Decision Trees, Decision Trees (PART), PRISM, Kernel Density, Instance Based1, Decision Tables.

We design 3 experiments and provide the summarized results here (the details will be provided upon request).

Experiment 1. We compare the use of dual ensembles ($E_{l=yes}$ and $E_{l=no}$) with that of single ensembles (either $E_{l=yes}$ or $E_{l=no}$). The average UCincrease (increase in uncertain regions) is almost 53% and the average PRgain is -1.04%. It is observed that in general, dual ensembles are not only more accurate, but also separate certain and uncertain instances better than single ensembles.

Experiment 2. We compare the learned dual ensembles to 10 randomly selected dual ensembles. Each classifier is randomly chosen and learns from the training image. Although the average PRgain is only decreased by 2.01%, the UCincrease increases significantly by 846.6%. Hence, it is necessary to search for relevant dual ensembles, as random dual ensembles work poorly in reducing UC.

Experiment 3. We compare the learned dual ensembles to results obtained by the classification rules of domain experts. The experts' rules outperform the learned dual ensembles in terms of PRgain by 14.6%, but the number of uncertain instances (UC) increases by 408.4% The high PRgain and high UC for the expert classification rules is due to the fact that an expert can only directly work on the former (designing highly general and accurate rules), but not on the latter (finding low UC rules). Our system is particularly designed to compensate in this shortcoming.

5 Conclusion

We present a novel approach to active learning with ensembles of classifiers. One ensemble is trained for *each class*. Extensive experiments were conducted in the real-world domain of detecting seaweed in aerial images.

References

[Agrawal and Srikant, 1994] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. VLDB*, pages 487–499, 1994.

[Witten and Frank, 2000] I. Witten and E. Frank. *Data Mining: Practical Machine Learning tools and techniques with java implementations*. Morgan Kauffmann, 2000.